

Abstract

There is provided a novel approach for generating multilingual text-to-phoneme mappings for use in multilingual speech recognition systems. The multilingual mappings are based on the weighted output from a neural network text-to-phoneme model, trained on data mixed from several languages. The multilingual mappings used together with a branched grammar decoding scheme is able to capture both inter- and intra-language pronunciation variations which is ideal for multilingual speaker independent recognition systems. A significant improvement in overall system performance is obtained for a multilingual speaker independent name dialing task when applying multilingual instead of language dependent text-to-phoneme mapping.